Molecular Biology and Biotechnology

A Comprehensive Desk Reference

Edited by
Robert A. Meyers

Robert A. Meyers, Ph.D.
3715 Gleneagles Drive
Tarzana, CA 91356, USA

Management Supervised by:
Chernow Editorial Services, Inc.,
1133 Broadway, Suite 721, New York, NY, USA

Cover design by: G & H SOHO, Inc.

Cover art courtesy of Dr. Alexander Wlodawer from Figure 1 of his article,
"AIDS, Inhibitor Complexes of HIV-1 Protease in."
Art prepared by Dr. Jacek Lubkowski.

Library of Congress Cataloging-in-Publication Data

Molecular biology and biotechnology : a comprehensive desk reference /
Robert A. Meyers, editor.
p. cm.
Includes bibliographical references and index.
ISBN 1-56081-569-8 (alk. paper). — ISBN 1-56081-925-1 (pbk.: alk. paper)
1. Molecular biology—Encyclopedias. 2. Biotechnology—Encyclopedias.
I. Meyers, Robert A. (Robert Allen), 1936-
QH506.M66155 1995
574.8'8'03—dc20 95-9063
CIP

Printed in the United States of America

ISBN 1-56081-569-8 (hardcover)

Printing History:
10 9 8 7 6 5 4 3 2 1

ISBN 1-56081-925-1 (softcover)

Printing History:
10 9 8 7 6 5 4 3 2 1

Published jointly by:

VCH Publishers, Inc.	VCH Verlagsgesellschaft mbH	VCH Publishers (UK) Ltd.
220 East 23rd Street	P.O. Box 10 11 61	8 Wellington Court
New York, NY 10010	D-6940 Weinheim	Cambridge CB1 1HW
USA	Federal Republic of Germany	United Kingdom

Fax: (212) 481-0897
E-mail address: order@vch.com

binding site is often included in the vector just preceding position at which the protein coding sequence is to be placed. leotide sequences frequently placed in vectors for their efficient me sites include chemically synthesized optimal consensus nces, the T7 g10-L region, or those of other highly translated While these standard translation initiation systems are suit- for adequate expression of most proteins, the sequence in rea has to be optimized to gain the highest level of protein uction. Translational limitations can also occur as a result e specific nature of the protein coding region (e.g., unusual position, unique codon usage, or special secondary structures). se problems cannot be addressed at the expression vector level.

Restriction Sites for Cloning

eful selection of the location of the restriction endonuclease avage site at which the protein coding sequence is to be intro- ed into the plasmid can allow the expression of the protein in most useful form. In many cases it is desired to produce the plete intact natural protein, beginning with its N-terminal no acid and ending with the native C-terminal amino acid. This quently is the case when the exact structure of a mature, natural, rmaceutically active protein is needed. To position the coding uence correctly with respect to the ribosome binding site, a cific restriction endonuclease site incorporating an ATG se- nce is used to place the translation initiation codon appropriately the vector. Two commonly used sites that contain an ATG and ly exist in any given DNA sequence are *Nco* I and *Nde* I. A ood unique restriction site is often used where the C-terminal of the gene is joined to the vector to orient the protein coding ment during cloning. Such a construct would then enable transla- of the complete cloned protein coding region.

Although not all mature proteins have an N-terminal methionine, is the primary product formed in bacterial systems. The extra no acid can be removed from the protein after synthesis by in o methods, or the desired protein can initially be made as part longer peptide chain with subsequent cleavage by a specific ease to generate the N-terminus of the mature protein product. riety of N-terminal protein carrier elements have been used ake protein fusion products. In the formation of a fusion ein, the vector carries the appropriate signals for translation ation, as well as a coding region of the N-terminal component protein and a restriction endonuclease cleavage site where DNA fragment bearing the foreign gene can be placed in the per reading frame. The properties of the N-terminal component protein fusion can be used to assist in purification of the on protein. For example, if this segment encodes an easily ified protein such as the maltose binding protein or glutathione nsferase, the combined fusion product can be readily isolated g affinity chromatography. The addition of tracts of basic, ic, or metal-binding amino acids in the carrier protein can also used to aid purification. This protein fusion concept can be ized to allow for protein processing in vivo. For example, if N-terminal segment encodes a localization signal, the protein be directed to the periplasmic space of *E. coli* or, in some s, into the extracellular medium. This type of export can aid in purification and stabilization of the protein. A vector employing otein A as the carrier is one of this type.

In some applications the expression of the entire protein is not essary. For raising antibodies or for the detection of an antigen by antibodies, only a recognizable epitope is necessary, so in these cases expression of only a short protein segment is sufficient. The cloned segment comprising the antigenic site can be attached to a protein carrier segment that will allow effective exposure of the foreign peptide. Placement of the foreign segment at the C-terminal end of β-galactosidase has been used effectively for this purpose. In some cases a suitable surface protein can be used as the carrier portion to ensure that the foreign protein segment is localized to the surface of a cell or bacteriophage particle, thus presenting the immunologically active antigen or antibody species in a way that allows detection, isolation, or use of the living cell or virus particle.

3 HOST CELL FEATURES

3.1 *Escherichia coli*

The use of the correct host can have significant impact on the final yield of the desired protein product. Factors generally useful to the host include the ability to grow rapidly and to a high density, the ability to be transformed in a manner that facilitates the introduction of the DNA construct, and the possession of a low recombination and mutagenic rate, to ensure that the plasmid is not frequently lost, damaged, or otherwise inactivated. Stability to degradation is a major problem in the production of certain proteins. Therefore, host cells with reduced protease levels have been used to enhance the in vivo stability of the foreign protein of interest. Commonly used mutations in *E. coli* that have reduced protein degradation rates because of the inactivation of proteases are *lon, rpo H,* and *clp.* In some cases the protein stability can be addressed at the expression vector level by using a fusion protein construct. The presence of the longer carrier protein often will effectively stabilize a small foreign peptide segment. Translation of large proteins is also a limitation in many strains; however other *E. coli* strains may have an increased ability to translate the particular product.

3.2 Other Hosts

While *E. coli* continues to be the most widespread expression system, other host–vector systems have certain advantages. The strengths of the *E. coli* system are the variety of vectors and specifically altered hosts available and the well-studied methods for manipulating this organism. High levels of production can be attained. However protease problems, formation of inactive inclusion bodies containing the product, and the lack of a eukaryotic glycosylation system limit production of a number of proteins from mammalian sources. Other bacteria have received some attention owing to their ability to grow on particular compounds, their potential for secretion of the protein, or their industrial potential.

Yeasts are a suitable production system for a number of processes. Not only are they well studied like *E. coli,* and amenable to scale-up, but they are able to carry out some posttranslational modification of eukaryotic proteins.

Fungi can produce high yields of commercial proteins and are reasonably capable of glycosylation and secretion of proteins into the medium. Methods have not been so completely developed, and the use of organisms of this class has not been widespread.

Viruses that infect insect cells (baculoviruses) have gained attention as a system for producing glycosylated proteins at reasonably high levels but at less expense than is incurred using mammalian cells. The vaccinia virus has been used to express foreign antigens in whole animals, demonstrating its potential for use as a vaccine

vector system. Although mammalian cells have been studied extensively and have the advantage of producing truly identical processed mammalian protein, the costly, difficult scale-up has limited commercial production with this system.

4 PERSPECTIVES

The ability to clone and express high levels of proteins in other organisms has led to great advances in the speed and detail with which biological systems can be analyzed. With the expansion of this technology to other organisms and the construction of more complicated and sophisticated derivatives of currently studied systems, the impact of this area will increase. Recombinant protein production has now begun to see use in applications to industrial microorganisms, the formation and use of transgenic plants and animals, and the analysis of unknown proteins associated with genetic disorders.

See also DNA MARKERS, CLONED; *E. COLI* GENOME; FUNGAL BIOTECHNOLOGY; GENE EXPRESSION, REGULATION OF; PLASMIDS; YEAST GENETICS.

Bibliography

Alitalo, K. K., Huhtala, M.-L., Knowles, J., and Vaheri, A., Eds. (1990) *Recombinant Systems in Protein Expression.* Elsevier, Amsterdam.

Barr, P. J., Brake, A. J., and Valenzuela, P., Eds. (1989) *Yeast Genetic Engineering.* Butterworths, Boston.

Goeddel, D. V., Ed. (1990) *Gene Expression Technology,* Vol. 185 in *Methods in Enzymology.* Academic Press, San Diego, CA.

Henninghausen, L., Ruiz, L., and Wall, R. (1990) Transgenic animals—Production of foreign proteins in milk. *Curr. Opin. Biotechnol.* 1:74–78.

Hruby, D. E. (1990) Vaccinia virus vectors: New strategies for producing recombinant vaccines. *Clin. Microbiol. Rev.* 3:153–170.

Kriegler, M. (1991) *Gene Transfer and Expression: A Laboratory Manual.* Freeman, New York.

Old, R. W., and Primrose, S. B. (1990) *Principles of Gene Manipulation: An Introduction to Genetic Engineering,* 4th ed. Blackwell Scientific Publications, Oxford.

O'Reilly, D. R., Luckow, V. A., and Miller, L. K. (1992) *Baculovirus Expression Vectors—A Laboratory Manual.* Freeman, New York.

Reznikoff, W., and Gold, L., Eds. (1986) *Maximizing Gene Expression.* Butterworths, Boston.

Ridgeway, A. A. (1988) Mammalian expression vectors. *Biotechniques* 10:467–492.

Vasil, I. L. (1990) The realities and challenges of plant biotechnology. *Bio/Technology* 8:296–301.

Wang, L.-F., and Doi, R. H. (1992) Heterologous gene expression in *Bacillus subtilis.* In *Biology of Bacilli: Applications to Industry,* R. H. Doi and M. McGloughlin, Eds., pp. 63–104. Butterworth-Heinemann, Boston.

EXTRACELLULAR MATRIX

Linda J. Sandell

Key Words

Exon DNA Of the gene that is represented in the mature RNA product.

Extracellular Matrix Material lying adjacent to and bet cells.

Gene Structure Organization of the gene, including the pro exons, and introns.

Intron DNA of the gene that is removed from the m RNA product.

mRNA Splicing The removal of introns from the pre-m and joining of exons in mature RNA; thus introns are sp out, while exons are spliced together.

The extracellular matrix is necessary for development and nor functioning of all cell types in an organism: it is made up of sec proteins and glycoproteins whose complex interactions deter the matrix properties. The composition of the extracellular m is controlled at the level of gene expression by providing quantity of mRNA sufficient to produce adequate proteins and alternative splicing of pre-mRNA to generate proteins with proper functional domains.

1 INTRODUCTION

The extracellular matrix (ECM) of all tissues is a complex mix of secreted proteins that collectively play a critical role in deter ing and maintaining tissue function. The ECM is located prim around connective tissue cells and under epithelia. ECM pro range from the multifunctional fibronectins and thrombospo to the large families of collagen types, proteoglycans, and lami among others. Fibrillar collagens types I, II, and III are the prin ECM proteins that confer the structural characteristics typic tissues such as bone, skin, blood vessels, and cartilage. Other fu proteins such as the fibronectins, laminins, and tenascin play crit roles in cell–cell interactions, cell migration, and cytoskeletal nization. The structure and function of extracellular matrix pro have been reviewed recently in two very informative books Hay (1993) and Kreis and Vale (1993). While not inclusive contribution discusses the molecular biology of the major categ of ECM components.

1.1 REGULATION OF ECM GENE EXPRESSION

It is becoming increasingly clear that the regulation of ECM expression, both transcriptionally (to regulate quantity) and p transcriptionally (using alternative splicing of mRNA to inc or remove functional domains), is of crucial importance in morphogenesis and cell differentiation, cell migration and proli tion, wound healing, and disease processes such as fibrosis arthroses. ECM expression is regulated by a wide variety of gr factors and cytokines, being generally stimulated by, for exam transforming growth factor beta, and insulin-like growth fa I and II, and inhibited by interleukin 1 and interferon ga Interestingly, as more information accumulates regarding re tion of ECM expression, it is apparent that ECM molecules are independently controlled in their expression and that regulati dependent on cell type. Table 1 shows some examples of regul of ECM molecules by cytokines. This is a rapidly emerging and more information is published every day. It is now known certain ECM molecules can effect the expression of themselve